PB89-197719

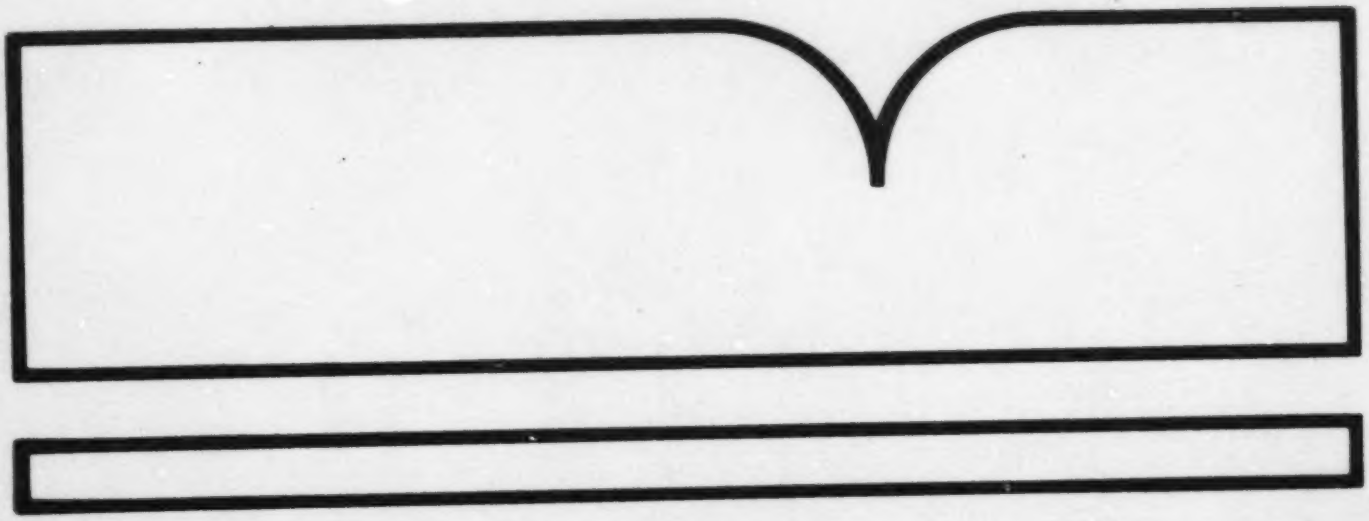
Methods Used for Managing and Analyzing the Data Stored in the NCLAN Data Library

North Carolina State Univ. at Raleigh

Prepared for:

Corvallis Environmental Research Lab., OR

1988



U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/J-88/274	2.
3. RE PB89-197719	
4. TITLE AND SUBTITLE Methods used for managing and analyzing the data stored in the NCLAN data library	5. REPORT DATE 1988
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Susan E. Spruill, Virginia M. Lesser, Matthew C. Somerville and John O. Rawlings	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS North Carolina State U.	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO. IAG DW12931347
12. SPONSORING AGENCY NAME AND ADDRESS US Environmental Protection Agency Environmental Research Laboratory 200 SW 35th Street Corvallis, OR 97333	13. TYPE OF REPORT AND PERIOD COVERED Journal Article
	14. SPONSORING AGENCY CODE EPA/600/2

15. SUPPLEMENTARY NOTES

Environmental Pollution 53:462-465. 1988.

16. ABSTRACT

The final library file of formatted data contains nearly 250 individual NCLAN data sets comprised from 43 studies from four continental regions collected over 7 years (1980-1986). Data collected from each study were formatted into three categories: 1) descriptive information, 2) air quality data, and 3) plant-response data. Coordinated analysis of the data included developing dose-response functions for the economic assessment incorporating experiment to experiment variation into measures of precision of estimated relative yield losses and investigating alternative methods of analyses of unbalanced data that properly took into account correlated errors common to many individual experiments and to data combined over random experiments. Dose-response models were developed for economic assessment by combining all data for each crop species to model assumptions and homogeneity of response to ozone over cultivars, sites and years. Methods used in these combined analysis included analyses of variance, polynomial and nonlinear regression analyses. Transformations of the data and of models were used when necessary based on development of residual and normal plots, Box-Cox analyses and Barlett's Chi-square Test for homogeneity of variance. An 'average' response was computed from models for each species, even though the response was known to be heterogeneous in most cases.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT Release to Public	19. SECURITY CLASS (This Report) Unclassified	21. NO. OF PAGES 6
	20. SECURITY CLASS (This page) Unclassified	22. PRICE PC A02/MF A01 6.95

EPA Form 2220-1 (9-73)

TABLE 1

Summary of Years in Which NCLAN Studies Were Conducted by Species and Site

Crops	*Sites*						
	Argonne, IL	*Beltsville, MD*	*Ithaca, NY*	*Raleigh, NC*	*Riverside, CA*	*Shafter, CA*	*Tracy, CA*
Alfalfa	1984				1984		
Barley						1982	
						1983	
Clover-Grass*			1984	1984			
			1985	1985			
Corn	1981						
	1985						
Cotton						1981	
				1982		1982	
				1985	1985		
Kidney Bean			1980				
			1982				
Lettuce						1983	
Peanut				1980			
Sorghum	1982						

(continued)

RTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

Methods Used for Managing and Analyzing the Data Stored in the NCLAN Data Library

Susan E. Spruill, Virginia M. Lesser, Matthew C. Somerville and John O. Rawlings

North Carolina State University, Raleigh, North Carolina, USA

Two major objectives of the NCLAN Data Synthesis and Analysis Project are: (1) to collect, format, and verify all NCLAN data in a central library (NCLAN Data Library) to facilitate coordinated analyses of all data and to

TABLE 1—*contd.*

Crops	Sites						
	Argonne, IL	*Beltsville, MD*	*Ithaca, NY*	*Raleigh, NC*	*Riverside, CA*	*Shafter, CA*	*Tracy, CA*
Soybean	1980						
		1981	1981	1981			
		1982		1982			
	1983[b]	1983		1983			
				1984			
	1985						
	1986			1986			
				1983			
Tobacco							1981
Tomato							1982
Turnip				1980			
Wheat	1982		1982				
	1983		1983				

[a] Ithaca used Red clover–timothy mix; Raleigh used Ladino clover–fescue mix.
[b] Argonne had two soybean studies in 1983.

ease access to NCLAN data for valid research objectives; and (2) to coordinate analysis of all NCLAN dose-response data. These analyses include developing dose-response functions for the economic assessment, incorporating experiment-to-experiment variation into measures of precision of estimated relative yield losses, and investigating alternative methods of analyses of unbalanced data that properly take into account correlated errors common to many individual experiments and to data combined over random environments.

The data collected for each study of the NCLAN program were formatted into three categories in the Data Library: (1) descriptive information, (2) air quality data, and (3) plant-response data. Quality assurance checks were employed on each data set in the data library to verify accuracy of data transfer from each site to the Library.

The final Library file of formatted data contains nearly 250 individual NCLAN data sets and utilizes approximately 20 megabytes of space. Forty-three studies from four continental regions (using a total of eight field sites) collected over 7 years (1980–1986) are represented (Table 1). The first objective of the project was completed in early 1987 with the inclusion of data from the 1986 studies from Argonne and Raleigh.

For the second objective, dose-response models were developed for economic assessment by combining all data for each crop species to test model assumptions and homogeneity of response to O_3 with respect to cultivars, sites, and years. Methods used in these combined analyses included analyses of variance, polynomial and nonlinear regression analyses; development of residual and normal plots; Box–Cox analyses (Box & Cox, 1964; Bartlett's Chi-square Test for homogeneity of variance (Steele & Torrie, 1980); and other statistical tools to best interpret these data. Transformations of the data and of the models were used when necessary to better satisfy the basic assumptions (Bartlett, 1947). Weighted least squares was used when heterogeneous variances could not be corrected. Common dose-response equations were computed for all subsets showing homogeneity of response. Models, including all data for a crop species, were also computed to provide an 'average' response for each species, even though the response was known to be heterogeneous in most cases.

References

Bartlett, M. S. (1947). The use of transformations. *Biometrics*, 3, 39–52.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *J. Royal Statist. Soc.* B-26, 221–43, Discussion 244–52.

Steele, Robert G. D. and James H. Torrie (1980). *Principles and procedures of statistics: A biometric approach*. 2nd edn McGraw-Hill Book Company, New York, NY.

TABLE 1—contd.

Crops	Sites							
	Argonne, IL	*Beltsville, MD*	*Ithaca, NY*	*Raleigh, NC*	*Riverside, CA*	*Shafter, CA*	*Tract, CA*
Soybean	1980			1981			
	1983[a]	1982		1982			
		1983	1981	1983			
	1986			1984			
	1985						
	1986			1986			
Tobacco				1983			
Tomato							1981
							1982
Turnip				1980			
Wheat	1982		1982				
	1983		1983				

[a] Ithaca used Red clover–timothy mix; Raleigh used Ladino clover–fescue mix.
[b] Argonne had two soybean studies in 1983.

ease access to NCLAN data for valid research objectives; and (2) to coordinate analysis of all NCLAN dose-response data. These analyses include developing dose-response functions for the economic assessment, incorporating experiment-to-experiment variation into measures of precision of estimated relative yield losses, and investigating alternative methods of analyses of unbalanced data that properly take into account correlated errors common to many individual experiments and to data combined over random environments.

The data collected for each study of the NCLAN program were formatted into three categories in the Data Library: (1) descriptive information, (2) air quality data, and (3) plant-response data. Quality assurance checks were employed on each data set in the data library to verify accuracy of data transfer from each site to the Library.

The final Library file of formatted data contains nearly 250 individual NCLAN data sets and utilizes approximately 20 megabytes of space. Forty-three studies from four continental regions (using a total of eight field sites) collected over 7 years (1980–1986) are represented (Table 1). The first objective of the project was completed in early 1987 with the inclusion of data from the 1986 studies from Argonne and Raleigh.

For the second objective, dose-response models were developed for economic assessment by combining all data for each crop species to test model assumptions and homogeneity of response to O_3 with respect to cultivars, sites, and years. Methods used in these combined analyses included analyses of variance, polynomial and nonlinear regression analyses: development of residual and normal plots; Box–Cox analyses (Box & Cox, 1964; Bartlett's Chi-square Test for homogeneity of variance (Steele & Torrie, 1980); and other statistical tools to best interpret these data. Transformations of the data and of the models were used when necessary to better satisfy the basic assumptions (Bartlett, 1947). Weighted least squares was used when heterogeneous variances could not be corrected. Common dose-response equations were computed for all subsets showing homogeneity of response. Models, including all data for a crop species, were also computed to provide an 'average' response for each species, even though the response was known to be heterogeneous in most cases.

References

Bartlett, M. S. (1947). The use of transformations. *Biometrics*, **3**, 39–52.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *J. Royal Statist. Soc.*, B-26, 221–43, Discussion 244–52.

Steele, Robert G. D. and James H. Torrie (1980). *Principles and procedures of statistics: A biometric approach*. 2nd edn McGraw-Hill Book Company, New York, NY.

EPA/600/J-88/274

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.

Methods Used for Managing and Analyzing the Data Stored in the NCLAN Data Library

Susan E. Spruill, Virginia M. Lesser, Matthew C. Somerville and John O. Rawlings

North Carolina State University, Raleigh, North Carolina, USA

Two major objectives of the NCLAN Data Synthesis and Analysis Project are: (1) to collect, format, and verify all NCLAN data in a central library (NCLAN Data Library) to facilitate coordinated analyses of all data and to

TABLE 1
Summary of Years in Which NCLAN Studies Were Conducted by Species and Site

Crops	*Sites*						
	Argonne, IL	*Beltsville, MD*	*Ithaca, NY*	*Raleigh, NC*	*Riverside, CA*	*Shafter, CA*	*Tracy, CA*
Alfalfa	1984				1984		
					1985		
Barley						1982	
						1983	
Clover-Grass[a]			1984	1984			
			1985	1985			
Corn	1981						
Cotton	1985						
						1981	
				1982		1982	
				1985	1985		
Kidney Bean			1980				
			1982				
Lettuce						1983	
Peanut				1980			
Sorghum	1982						

(continued)